

Unlocking Biological Insights Through AI-driven Multi-Omics Integration.

Shivaprasad Patil

Associate Director, Predictive AI & Data, AstraZeneca

Abstract

Background: Multi-omics data integration is critical for understanding complex biological systems, but traditional approaches face scalability and interpretability challenges.

Methods: We present two complementary frameworks: • MultiOmicsBind - A deep learning framework using ImageBind-inspired binding modality for efficient $O(n)$ complexity multi-modal contrastive learning • OmicsNet - A network-based approach for visualizing and analyzing multi-omics relationships.

Results: MultiOmicsBind successfully integrates static + temporal multi-omics profiles to capture dynamic molecular changes during biological processes to reveal disease progression.

OmicsNet maps multi-omics regulatory networks and uncovers conserved functional modules that coordinate cellular responses across transcriptomic, proteomic, metabolomic and imaging modalities.

Conclusions: These frameworks provide scalable, interpretable solutions for multi-omics integration, enabling both hypothesis-driven and data-driven discovery in systems biology.

Introduction

Modern biological research generates diverse data types:

- Genomics: DNA variants, SNPs (100K-500K features)
- Transcriptomics: Gene expression (10K-50K genes)
- Proteomics: Protein abundance (5K-20K proteins)
- Metabolomics: Metabolite levels (1K-5K metabolites)
- Cell Painting: Morphological features (1K-5K features)

Current Limitations:

- Computational Complexity: Traditional all-pairs approaches scale as $O(n^2)$.
- Interpretability: Black-box models lack biological insights.
- Missing Data: Inability to handle incomplete modalities.
- Integration: Limited unified framework for analysis and visualization.

Our Solution

Two Complementary Approaches:

- MultiOmicsBind: Efficient deep learning with binding modality.
- OmicsNet: Interactive network-based visualization and analysis.



Explore the Repositories

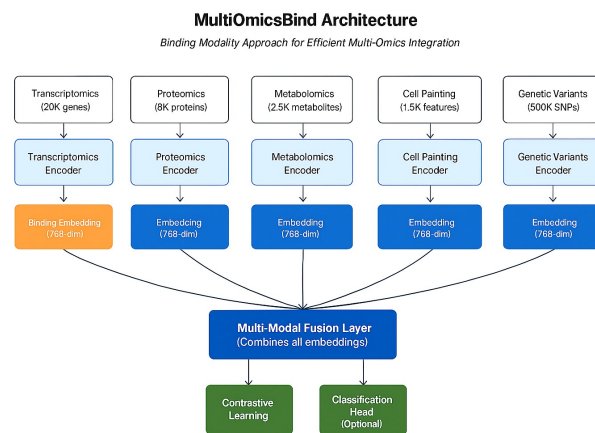
<https://github.com/shivaprasad-patil/MultiOmicsBind>

<https://github.com/shivaprasad-patil/OmicsNet>

Open for collaborations!

If you're interested in multi-omics, AI or drug discovery, feel free to reach out or contribute!

MultiOmicsBind



Key Innovation: Binding Modality Approach Biologically-Inspired Multi-Omics Integration

Core Biological Concept:

The binding modality acts as a molecular "anchor" that coordinates information flow between different biological layers, similar to how the transcriptome serves as a central information hub between genomic regulation and protein function.

Multi-Omics Integration:

Genomics → Transcriptomics (Binding) → Proteomics → Metabolomics → Imaging

Biological Interpretability

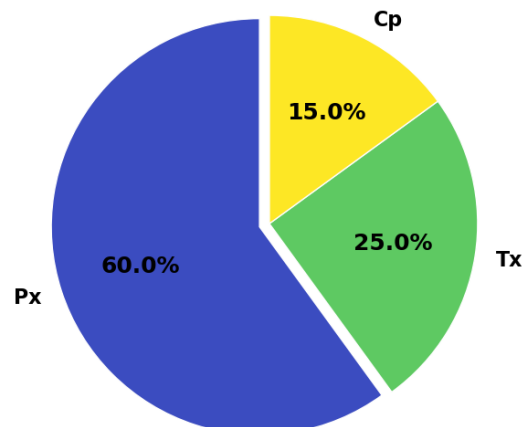
Feature Importance Analysis:

- Identifies key genes, proteins, metabolites, morphological features driving predictions.
- Maps biomarkers to known biological pathways
- Reveals novel therapeutic targets.

Cross-Modal Similarity:

- Measures coordination between omics layers.
- Identifies synchronized molecular profiles.
- Integrate static and temporal modalities.
- Combines embeddings for downstream analysis.

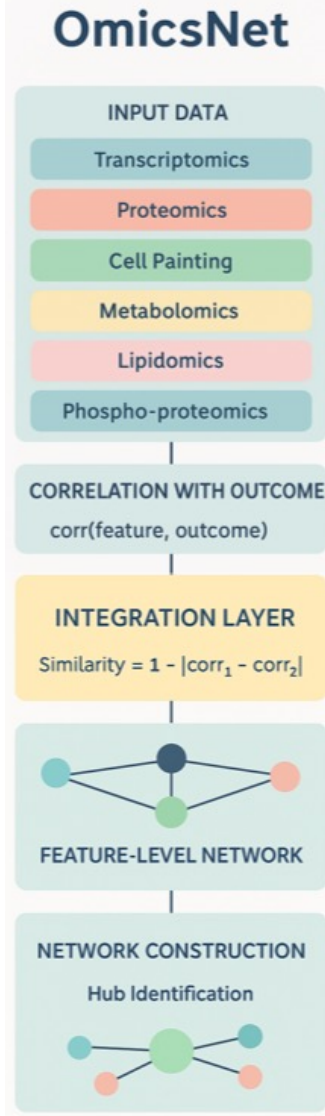
Modality Contributions



Test performance on simulated data: (240 samples) on drug response prediction using temporal multi-omics integration: Transcriptomics -Tx (6K genes) + Proteomics - Px (4K proteins × 5 timepoints) + Cell Painting - Cp (1.5K features).

Key Insight: Temporal proteomics data contributes the most to model predictions, highlighting the importance of dynamic molecular changes. The model effectively captures these temporal patterns, demonstrating the value of integrating time-resolved multi-omics data.

Omicsnet



Python package for integrating data from multiple omics modalities (transcriptomics, proteomics, cell painting, metabolomics, etc.) to build and analyze feature-level networks based on association patterns with outcome variables.

OmicsNet uses **feature-level integration** when samples don't overlap.

This approach identifies functionally related features across modalities even when measured in completely different samples.

Key Features

- Feature-Level Integration:** Integrate omics datasets with zero sample overlap using association-based similarity.
- Flexible Outcome Types:** Supports binary, continuous, and multinomial outcome variables with appropriate statistical methods.
- Feature Selection Methods:** Multiple approaches to identify top features (association, variance, network centrality, combined scoring)
- Network Construction:** Build integrated networks based on cross-modal feature similarities.

Conclusions

MultiOmicsBind: Integrates static + temporal multi-omics data, robustly handling missing values to create unified, interpretable embeddings for accurate prediction and biomarker discovery.

OmicsNet: Builds multi-layer biological networks, enabling visualization and analysis of cross-omics relationships and functional modules for biological interpretation.

Complementary Integration:

Combines deep learning-based unified embeddings from MultiOmicsBind with OmicsNet's interactive network analysis, bridging prediction and interpretation to enable pathway enrichment, hub gene identification, and hypothesis generation for novel biological insights.

References

- R Giridhar et al. 2023. CVPR.
- R Argelaguet et al. 2018. Mol Syst Biol.
- A Paszke et al. 2019. NeurIPS.
- S Adewale et al. 2023. arXiv.
- A Vaswani et al. 2017. arXiv.